图书情报工作
第 63 卷 第 10 期 2019 年 5 月

网络舆情场中观点簇丛的情感极化度测算*

■ 高俊峰[1] 黄微[2]

[1] 北华大学图书馆 吉林 132013 [2] 吉林大学管理学院 长春 132013

摘要：[目的/意义]提出网络舆情场内观点簇丛的情感极化度测算方法，为量化舆情受众情感态势和识别极化群体提供依据。[方法/过程]首先明确舆情受众情感极化的条件，再通过设置条件阈值筛选出满足条件的观点簇丛，在此基础上引入3个极化指标(受众吸引率、极端受众增长率、极化情感增长率)描述观点簇丛在测度时间窗口内的情感发酵程度。最后利用平滑权值，对观点簇丛在不同测度阶段的情感表现进行加权综合，得出其整体情感极化程度值。[结果/结论]观点簇丛在每个时间窗口的情感表现能为阶段性的舆情受众情感极化干预提供判断依据，而综合的情感极化度有助于准确识别场域内的敏感话题及群体，便于网络舆情的精准管控。

关键词：网络舆情治理 情感极化测度 观点分析 网络舆情场

分类号：G206.3

DOI:10.13266/j.issn.0252－3116.2019.10.012

1 研究背景

伴随着网络覆盖面的不断扩大，网络舆情已经成为公民利用网络表达诉求和参与社会管理的一种普遍现象。进入新时代以来，我国政府对网络舆情的重视程度日益提高，营造清朗网络空间正在成为建设中国特色社会主义互联网络的重要议题。网络的低成本、便捷性能够让不同身份背景的网络民众以灌水评论、立场投票等多种途径参与到社会热点事件中，在一定程度上为社会管理者提高管理决策水平提供重要参照。但与此同时，多元的价值观念和复杂的利益诉求也裹挟在网络舆情之中，使得网络舆情成为一把双刃剑，一旦应措失当，将会引发网络民众的情感极化乃至衍变成线上线下的突发事件。

从现实来看，网络舆情在向群体极化和网络突发事件的演化过程中，必然会依托若干个特别的网络空间或者平台，在此空间中围绕既定舆情事项，相近、相同的利益要求和立场倾向得到聚合，而不一致的观点态度被区分，这一舆情情感发酵平台即网络舆情场。

身处舆情场中的信息受众所发表的观点立场，其语义会有所指向(一般指向舆情事项中具体出现的人或事物)，笔者将被指向的实体定义为网络舆情事项的应事实体[1]。舆情受众的情感态度围绕引发舆情事项和推动舆情演化的应事实体汇集形成若干个观点簇丛，这种带有群体性力量的观点簇丛在影响舆情事项态势走向的同时，客观上也是民众对舆情事项背后问题的共识性反思，甚至是带有一定偏见的讨论，从而成为舆情参与者情感极化的温床。

对于由舆情话题和受众观点共同作用下所引发的群体情感极化现象，目前学术界已经从成因、识别和治理多个方面做出相应的研究：

(1)定性研究方面：有学者将网络群体性事件分为基于共同关注点的网络舆情和谣言、基于共同信念的网络舆论和网络暴力，以及基于共同目标的涉及线下的网络行动3类，为网络群体行为监管提供理论支撑[2]；有文献指出，基于一致性意见而形成的群体，其集群行为更倾向于该群体所持情感意见的实践化，并且对群体所采取的行为能够保持高度忠诚[3]；有研究更是直接指出群体性愤怒(group-based anger)是群体情感极化和群体行为的前置准备状态，而个体的情绪失控无法造成所属群体对相关情境或事件产生功能性反应[4]；有文献将网络舆论中的意见分歧和争论视作

＊本文系教育部人文社会科学研究青年基金项目"网络舆情场内信息受众的观点分析与情感极化管控研究"(项目编号:17YJC870006)研究成果之一。

作者简介：高俊峰(ORCID:0000－0003－0703－2606)，馆员，博士；黄微(ORCID: 0000－0002－7003－4631)，教授，博士生导师。

收稿日期:2018－09－13 修回日期:2019－01－08 本文起止页码:106－114 本文责任编辑:徐健

网民情感极化的驱动力,是话题网络拓扑结构发生改变的关键[5];有学者总结了网络集体行为意向的形成要素,并阐述生理测量法,自我报告法与量表测量法对意向预测上的差异,最后提出网络集体行为意向的形成周期和影响因素[6]。

(2)定量研究方面:有文献基于分位数回归模型,来制作网络群体极化预警的定量指标体系,并以网络论坛的真实数据为对象,对模型和指标的舆情应急可行性做出研究[7];有研究构建网络群体极化过程中政府-普通网民-意见领袖之间的博弈模型,并以不同阶段博弈主体的期望收益作为网络群体极化维稳的依据[8];有研究从健壮性的视角出发,考察网络舆情观点群落在遭受打击、核心端点被破坏的前提下,其群落在观点传播、情感动员和保持内部凝聚力 3 个方面的能力,以此作为判断观点群落是否具有极化趋向,是否必要采取干预措施的依据[9-10];还有文献通过定义舆情场内的受众个体观点交互规则和观点演化规则来仿真舆情极化的过程和预测群体极化的发生[11]。

大量研究均表明,监测和应对网民情感极化是网络舆情治理工作的核心,但目前所取得的研究成果在舆情受众情感极化的判断和测度问题上还不能达成一致性的结论。舆情受众情感极化的生发必然依托其隶属的观点阵营、簇丛,因而选择有效且便于获取的情感极化程度指标对舆情观点簇丛进行计算,是量化场域内不同舆情群体情感态势的合理方案。然而,并非舆情场中所有的观点簇丛都能呈现情感极化状态。这是因为不同舆情事项的应事实体性质具有差异,实体的行为类别和恶劣程度都会对受众的情感演化程度造成不同影响。因此,本研究直接将网络舆情场内出现的应事实体作为受众观点簇丛的划分依据,围绕某一应事实体的观点簇丛理论上能呈现出“积极、消极和中立”3 类,除“中立”类观点簇丛之外,其他两类簇丛可被认定为舆情受众生发情感极化的载体。但正如前文所言,并不是所有簇丛都一定会出现情感极化现象,所以本研究先根据情感极化定义设置若干个情感极化必要条件,对于场域内满足条件的簇丛加以筛选并认定其为待测度对象,再通过描述和计算对情感极化造成影响的要素状态(如受众个体情感强度、极端情感总量和簇丛规模等),得出某一时间阶段被测度观点簇丛的情感极化程度。而随着时间演变,各要素状态的变化也能够体现在下一时间阶段内测得的结果之中。最后引入指数平滑方法,对整个测度时间序列下的观点簇丛情感极化度进行综合计算,所得出的结果能够在一定程度上反映出某一立场阵营的情感极化表现,也能识别出舆情事项中易于引发极化情感的应事实体。进而为监测敏感议题、精准干预非理性舆情群体提供量化依据。

2 相关概念界定

2.1 网络舆情场内受众情感极化

早期提出的网络非理性言论形成机制中,将群体情感极化定义为群体内成员在发生观点交互之前即已持有某种立场倾向,并通过群体的商権讨论将其加强,使得一种观点或态度远超于原来的群体平均水平,最后变得极端和获得支配性地位[12]。

上述经典定义给出了极化的一般性质:立场一致性、情感极端性以及观点统驱性,前两个极化属性已经被广泛认可,但已然极化的群体所持观点到底是否必须压倒和支配一切其他群体?换而言之,在一个定域内,能否有一种以上的群体观点呈现出情感极化?笔者认为同时存在多个群体的情感极化现象是存在的。因为在网络环境中,“去身体化”与“去中心化”同时并存,一方面网络民众不必强制屈从于精英观点,另一方面网络民众的发言又没有现实中身份背景上的限制顾忌。这就导致竞争、对抗性的观点只要能获得一定认同,即会受到鼓励而非约束,舆情受众在捍卫各自观点阵营的同时,或将产生多种立场“势均力敌”的局面,为多轨迹、多向度的情感极化发生提供可能。此外,上文曾提及网络舆情场内或将存在多个应事实体的情况,围绕某一舆情应事实体的受众观点集结成簇,宏观上固然可与其他实体的观点簇丛立场趋同。但微观而言,不同应事实体或者同一实体的不同话题兴奋度和社会敏感性是不可等量齐观的,这也或将导致网络舆情场内情感极化群体不唯一的现象出现。

2.2 网络舆情场内受众情感极化条件

根据上述定义,不难归纳出情感极化发生的条件:

(1)条件 1:围绕应事实体所形成的观点簇丛需要具备一定的规模。情感极化现象带有明显的群体参与特征。否则即便有意见领袖在此过程中起到情感动员和引领的作用,但倘若没有引发舆情受众响应,也无法酝酿出情感极化所需要的整体情感强度,所以情感极化现象出现与否首先受到待测度簇丛所吸纳的受众数量的制约。此外,极化观点簇丛的规模也是在场域内保持与其他极化簇丛之间竞争优势的必要条件。情感极化是网络民众参加和动员的结果,所以无论受众个体的观点蕴含的情感态度如何极端,都无法代表所属

团体,零星的极端态度、言论是无法被视作群体情感极化的。

（2）条件 2:观点簇丛内的受众个体情感强度需要呈现极端化趋势。根据定义,情感极化的发生必然伴随着持有观点显著的极端化倾向,这种显著极端无法简单的等同于大量受众所持有情感的简单叠加,必须是持有极端化情感的个体要在所属群体中占据到相当的比例。因为在一个大规模的群体中,即使受众个体均普遍持有温和的非中立立场,简单的情感强度累加后,也或将碾压场域内其他规模相对较小的群体(即使受众个体情感态度普遍极端)。这种情况是违背情感极化定义的,所以在确认极化观点簇丛规模的同时,仍需考量其中的极化个体占有率。

（3）条件 3:观点簇丛内极端情感总量需要到达相应的强度。观点簇丛中的极端情感总强度是前两个条件的具体体现。因为围绕应事实体所触发生成的极端情感总强度是在场域内与其他极化群体进行博弈的着力点。网络舆情场内受众情感极化的轨迹和向度也是通过不同极化群体之间极端情感强度的对比来完成。因此,群体内的极端情感总量也是一项重要测度指标。

极化条件的界定与采用,是为了定量辨别场域内出现的观点簇丛在初始测度时间窗口下是否发生了情感极化,进而实现对达到极化条件的观点簇丛的选择。

3　测度机理

测度的整体思路是“先判断出待测度簇丛,再提取簇丛的极化指标,而后对簇丛分阶段测算其情感极化度,最后根据阶段性结果得出综合的情感极化值。”在以往的研究中,群体规模、话题讨论量、群体交互频率[13]等指标均被用于关于群体极化的测算。上述指标的共同特点是易于获得,且都能在一定程度上反映出群体的情感活动状态。笔者沿袭此思路,除了紧密围绕前文条件生成极化指标之外,力求选择易于获得且能够反映簇丛情感状态的指标来满足测度工作。

判断待测度簇丛步骤可以直接设置一系列的阈值来筛选出满足条件的观点簇丛,而簇丛的极化指标虽然也可以直接使用条件所对应的取值(簇丛规模、簇丛极化受众量、簇丛极化情感量),但在测度初始阶段已然满足极化条件的观点簇丛,其自身的极化情感总量、极端受众人数或会随着时间的演进而展现出不同的变化,直接采用各项取值计算出的结果势必会弱化观点簇丛情感整体演进态势的描述力度,因此笔者选择 3 个比率性指标“极端受众增长率”“极化情感增长率”“受众吸引率”作为观点簇丛的极化程度指标。前两个指标能描述出待测度簇丛的自有情感状态演变趋势,而后一个指标能作为一个基数来描述整个场域内各观点簇丛的消长对比情况。

上述 3 个指标形成一个测度公式,协同完成对各观点簇丛情感极化度的阶段性测算,而后再对阶段性结果进行处理,综合得出一个观点簇丛情感极化度的取值。具体测度步骤如图 1 所示:

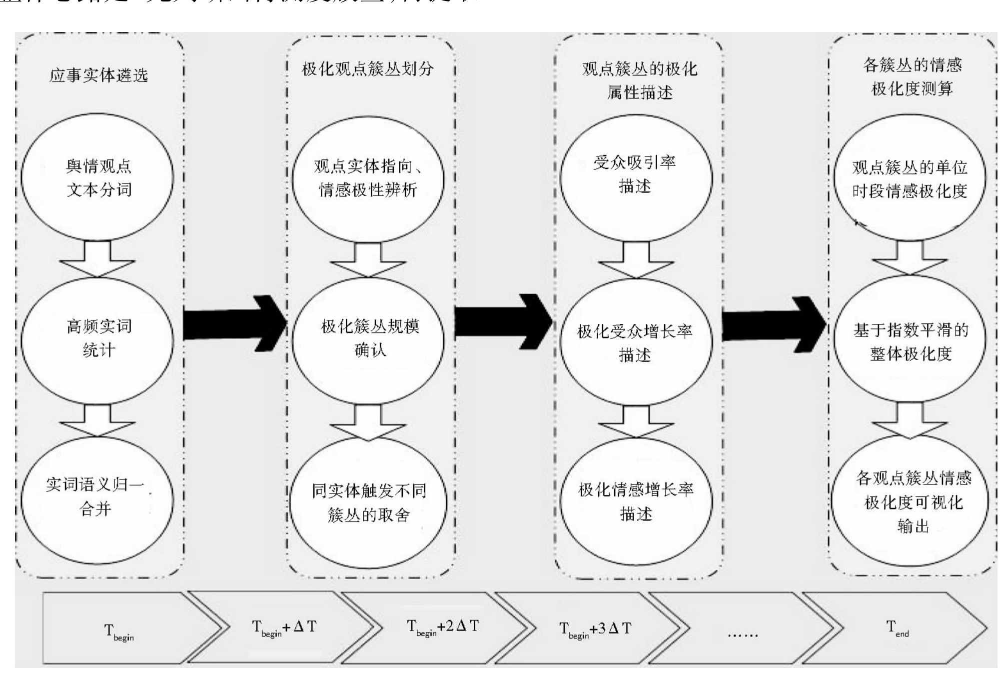


图 1　网络舆情场中观点簇丛情感极化度测算步骤

（1）确定观测时间序列。将整个测度时间段均分成 n 等份，使每个时间窗口宽度一致，而后从各个窗口内分析观点簇丛的指标变化。

（2）遴选应事实体。首先借助 NLPIR 汉语分词系统对舆情场内出现的观点文本进行分词，将文本中的实词进行拆解，得出其中的名词、动词以及修饰性的情感词汇，并利用“知网”等本体对词汇语义进行归一合并，摒除歧义。而后采用 TF-IDF 的词频统计思想，将满足频度阈值“Ω”的名词挑出，并根据上下文，人工补齐名词的舆情描述属性，以此凸显应事实体的舆情话题质性（同一应事实体或可具有多个话题，而受众对话题的态度又可被进一步细分，进而形成观点簇丛）。因为随着舆情事件的发酵和网民发散性认知，可能会出现新的应事实体和观点簇丛，所以此步骤需要在每个时间窗口下重复执行，下一步骤中的极化观点簇丛识别也需如此办理。

（3）极化观点簇丛识别。在词频统计步骤中，可以通过观点内容的语义分析，完成对场域内舆情观点的集结成簇。与此同时，必须摘选出不符合条件的情感中立型受众。通过将上文语义分析步骤得到的受众个体选择修饰实体的情感词汇进行极性和强度累加，能够将“应事实体＋话题”按接收到的显性观点逐条归纳为积极（Positive）、中立（Objective）、消极（Negative）三种立场之一（围绕同一应事实体话题形成的观点簇丛也与之对应，见图 2）。中立理性的观点无法形成极化情感，因此不作为测度对象。如此，或将出现围绕应事实体产生的观点簇丛以及情感极性不唯一的情况，所以需进一步观察实际存在的极化群体初始规模，设置当前群体人数阈值 Ant（观点簇丛必须在测度时段内拥有的受众总数不小于该值，则认定簇丛满足条件 1）、极化个体情感量阈值 Paet（个体观点中使用的情感词强度不小于该值，则认定簇丛满足条件 2）、极化情感总量阈值 Pet（簇丛中全部观点中使用的情感词强度累加后总值不小于该值，则认定簇丛满足条件 3）作为极化判断门槛，并设置同一应事实体的不同簇丛之间初始极化情感的对比阈值 Д（待测度簇丛与其他簇丛情感总量之比不能小于该值），对观点簇丛进行合理取舍。

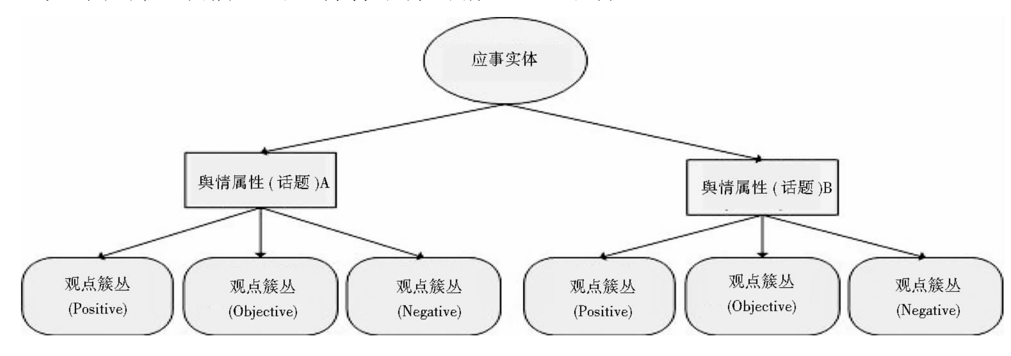


图 2　网络舆情场中观点簇丛遴选析出机制

（4）极化簇丛属性描述。对场域内待测度观点簇丛，需要从 3 个方面描述其极化特征：

受众吸引率（Audience attraction rate）＝（T_i 时段簇丛受众增量/T_i 时段场域内受众总增量）＊100%。

极端受众增长率（Growth rate of extreme audience）＝［（T_i 时段簇丛极化受众总量－$T_{(i-1)}$ 时段簇丛极化受众总量）/ $T_{(i-1)}$ 时段簇丛极化受众总量］＊100%。

极化情感增长率（Growth rate of polarized emotion）＝［（T_i 时段簇丛极化情感总量－$T_{(i-1)}$ 时段簇丛极化情感总量）/$T_{(i-1)}$ 时段簇丛极化情感总量］＊100%。

在极端受众增长率与极化情感增长率的测度工作中，或将出现 $T_{(i-1)}$ 时段簇丛极化情感总量、极化受众总量为零的情况，这种现象是由于新话题引入造成的，如果出现上述情况，则本研究默认 $T_{(i-1)}$ 时段簇丛极化情感总量、极化受众总量为场域内全体待测度观点簇丛的情感、受众总量的平均值，对于测度初始窗口的上述指标，则默认为 0。

待测度观点簇丛内受众的增量以新进受众是否在场域内发布文本型观点为基准，同一受众在不同观测窗口的观点发布行为不被视作受众增量，但如果其观点所含情感强度由温和向极端演进（根据前文的极化个体情感量阈值判断），则将其视作被触发情感极化的受众，纳入极端受众增长率和极化情感增长率的计算。

（5）各观点簇丛的阶段性情感极化度测算。上一环节求得的 3 组指标与舆情受众群体极化的 3 个条件相对应。又由于在极化簇丛识别步骤中已经剔除未能满足条件的观点簇丛，因此 3 组指标能直接勾勒出各观点簇丛的阶段性受众情感极化度。

首先,受众吸引率(Aar)用来描述场域内的各待测观点簇丛在时间序列下的规模增速。因为受众的显性立场一经抓取,即使下一时段出现立场反转,那么前一时刻该受众的极化状态仍然具有统计意义,所以指标 Aar 的值域为[0,1]。受众吸引率也能在一定程度上描述不同时态的舆情场内群体极化类型(单/双/多向度)。如前文所言,情感极化是群体参与的产物,个别的极端情感如果无法煽动、吸引足够多的受众,是无法被称为极化的。所以,一个观点簇丛同化、吸引受众的能力越强,则该簇丛内被触发极化情感的受众基数也越大。相应的,形成该观点簇丛的应事实体也越易于引起网民情感极化。就此而言,受众吸引率这一指标承担网络舆情场内各个观点簇丛之间情感极化度的横向比较功能。

此外,极端受众增长率(Grea)与极化情感增长率(Grpe)两个指标是从时间维度上纵向地描述待测观点簇丛极化条件的变化。理论上,极端受众人数以及簇丛内极化情感总量如果能保持一种稳步上升的态势,那么则认为此观点簇丛的情感极化度越高。因为上述两指标能直接展现观点簇丛的极化趋向,所以指标与情感极化度呈现正相关关系。与受众吸引率不同,极端受众增长率与极化情感增长率两个指标的取值区间为[0,+∞),二者都是描述观点簇丛内在的情感变化状态的,虽然测度数值能精准地反映出簇丛极化状态,但指标的全局描述能力较弱。而受众吸引率指标,虽然能从整个网络舆情场的角度考量各观点簇丛规模变化,但倘若单独用其刻画情感极化度,则精准性稍显不足。因此,综合考虑这 3 类指标是计算单位时间内观点簇丛的情感极化度的必然途径,为此笔者构建单位时段内的观点簇丛情感极化度(Emotion polarization degree)测度公式:

$$Epd(Tx)=\begin{cases}\delta Grea*Aar+\varepsilon Grpe*Aar\rightarrow\{Grea+Grpe\neq 0\}\\ 0\rightarrow\{Grea+Grpe=0\}\end{cases}$$

公式(1)

在公式(1)中,极端受众增长率(Grea)与极化情感增长率(Grpe)两个指标中倘若出现一方为 0,那么另一增长指标也必然为 0,此时不对公式进行运算,直接默认此情况下的观点簇丛在当前测度窗口内的情感极化度为 0。而后利用受众吸引率分别对两个增长率进行积运算,弥补指标单独刻画极化度上的局限,使得指标的统计意义均能发挥作用。此外“δ、ε”是研究设置的两个指标权重,以此描述极端受众增长率与极化情感增长率在展现情感极化度上的重要程度,根据前文分析得出的极化条件,笔者认为两类增长率在描述观点簇丛的极化能力上拥有相同的重要意义,因此将两权重值默认为 0.5。

从公式(1)中能够得出各测度时段(Tx)内观点簇丛的情感极化度,但研究最终需要得出一个具体的数值。考虑到原始时间序列中每个阶段性情感极化度对最终结果的影响程度均有不同,需要对各阶段性极化度赋予不同的权重,所以笔者选择指数平滑法对各观点簇丛的阶段性情感极化度进行综合计算,最终得出网络舆情场内观点簇丛的整体情感极化度。

设 β 代表权数,“OC”为待测度观点簇丛,n 为测度时间窗口个数,则根据权值选择机制,$\beta=2/n+1$,权数集合成等比数列,公比为 $\gamma=1-\beta$,每一个时间窗口对应唯一权数,对最终结果影响较大的最近时段极化度赋予最大权值[14],则观点簇丛“OC”的受众情感极化度计算公式为:

$$Epd(OC)=\frac{2}{n+1}Epd(T_n)+\left(\frac{n-1}{n+1}\right)\frac{2}{n+1}Epd(T_{n-1})+\cdots+\left(\frac{n-1}{n+1}\right)^{n-1}\frac{2}{n+1}Epd(T_1)$$

公式(2)

由于测度机制的限制,所以 T1 阶段的指标值为 0。

网络舆情场中观点簇丛的情感极化度的全部测度流程可形式化表述如下:

Input:舆情观点的应事实体词频统计阈值 Ω,观点簇丛人数阈值(Ant),极化个体情感量阈值(Paet),极化情感总量阈值(Pet),测度时序(T_{begin},$T_{end}/\Delta T$)同一应事实体的不同观点簇丛之间初始极化情感度的对比阈值 Л;

Output:各观测时间窗口内的待测度观点簇丛所对应的观点簇丛情感极化度 EPD(T_X),各观点簇丛的整体情感极化度 EPD(OC);①调用 NLPIR 汉语分词系统[15],对网络舆情场内的观点文本进行语义处理,调用词频统计阈值 Ω,获得舆情应事实体集合 Set1;②调用情感本体[16],获得舆情场内的受众立场指向,进而获得观点簇丛集合 Set2,统计得出每一观点簇丛的受众人数 An;③再次调用情感本体,对各观点簇丛进行极化情感总量(Pe)、极化受众情感量(Pae)计算,并调用极化个体情感量阈值“Paet”识别出当前时间窗口下每个观点簇丛中的极端受众与数量(Ean);④调用观点簇丛人数阈值(Ant),极化个体情感量阈值(Paet),初始极化情感度的对比阈值 Л,对集合 Set2 内同一实体附着的观点簇丛进行合理取舍,得出待测度观点簇

丛集合 Set3;⑤将 Set3 中的每一个观点簇丛表述成一个四元组(Pe,Pae,Ean,An),引入测度时间要素,计算出每个簇丛在当前时间窗口下的受众吸引率(Aar)、极端受众增长率(Grea)与极化情感增长率(Grpe);得出当前时段下的各观点簇丛情感极化度 Epd(T?/OC?);⑥更新时间窗口,重复执行步骤①－⑤,直到达到预置的终止测度时间节点,计算出平滑权数 β,观测各簇丛所出现的时间窗口序列以及对应的权数值,得到各观点簇丛的整体情感极化度 Epd(OC?)。

4 实证研究

笔者以近期发生的"高铁霸座"事件作为实证研究对象,以"新浪微博"中发起的此事件主题讨论区作为本次研究的网络舆情场,数据采集时段 2018 年 8 月 23 日 15:00－2018 年 8 月 31 日 15:00,T = 8,ΔT = 24h。数据清洗后,测度时段的场域内发布观点受众共计 16 456 人次。由于实证的目的是验证测度方法,因此不必对文中的阈值设置做过多要求。在研究过程中,笔者将各个初始判断阈值分别设置为:Ω = 0.1,Ant = 400,Paet = 20,Pet = 4 500,同一应事实体的不同立场群体间初始极化情感度的对比阈值 Д < 0.39。考虑到舆情场内的观点数据集中,语义等同于"霸座男"的词频过高,因此在统计中但凡出现"霸座男"以及相同语义的词汇时,直接统计其伴随舆情属性,即话题词汇。通过对高频话题词汇的语义消歧和语义人工合并,得到满足阈值条件的结果如下所示:

{OC1,霸座男 + 占座(3 493)}、{OC2,霸座男 + 道歉(2 942)}、{OC3,霸座男 + 论文抄袭(2 349)}、{OC4,霸座男 + 高学历(1 952)}、{OC5,霸座男 + 处罚(2 905)}、{OC6,霸座男 + 轮椅挑衅(2 815)}。

而"济南铁路局 + 调查结果不违法""霸座男 + 热衷相亲""网络暴民 + 人肉搜索""霸座男 + 社科院职工""中科院 + 躺枪"等应事实体的词频出现率都无法达到给定阈值,因此本研究最终得出 1 个应事实体的 6 个话题。

调取情感本体,识别受众个体文本观点中的情感极性。本研究选择的情感本体将情感词汇分为强度递进的五值("1,3,5,7,9"),词汇的褒贬义通过强度值的正负来勾勒[16]。在本研究中,受众观点中情感修饰词汇强度总值的"正/负"对应其立场的"Positive/Negative"。整个观点簇丛的情感极性取值由所含个体的情感叠加获得。此外,修饰情感词汇的程度副词的出现,往往会对当前情感词的强度值产生波动性影响。为了准确识别信息受众对应事实体的立场,本研究借助文献[17]提出的程度副词赋权方案,将程度副词分为"极、最(1.5)""很(1.3)""稍微(1.1)""适度(1)""不够(0.8)""过分(－0.5)""否定(－1)"7 类,而后与其修饰的情感词值进行乘积运算,最后累加出情感强度/极性。倘若同一受众观点中包含多个话题对象(隶属多个观点簇丛),则分别计算对不同话题对象的情感取值后按前文提出的规则进行归类判断。

根据情感本体计算出的各受众的情感极性和强度,剔除中立受众后进一步分析初现时间窗口下各应事实体 + 话题产生的观点簇丛的极化表现情况,各观点簇丛的极化情感极性以"Negative"为主流,除 OC6 的初现窗口为 T8 外,其余观点簇丛的初现窗口均为 T1,Ant、Pet 阈值达标。此外,对立立场(positive)簇丛的情感量与之对比结果均小于 0.2,满足对比阈值条件。得出各观点簇丛规模如图 3 所示:

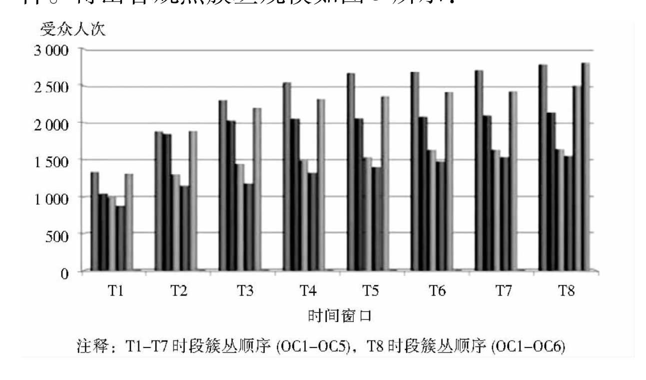


图 3 各测度时段下的观点簇丛的受众总量

通过前述步骤中的情感极性强度计算,得出实证数据在测度时序下各簇丛的极化受众体量以及极化情感总量如图 4 和图 5 所示:

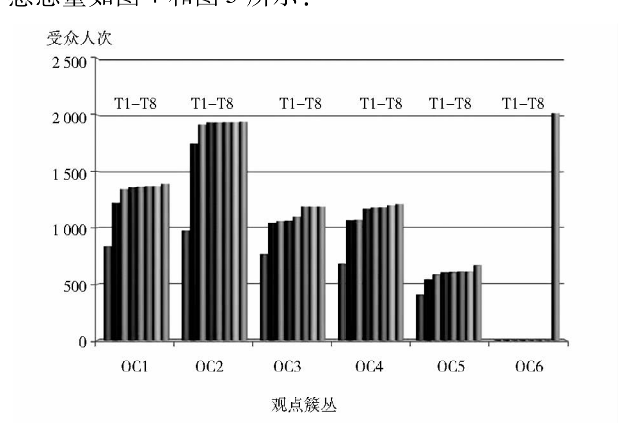


图 4 各测度时段下观点簇丛的极化受众总量

调用观点簇丛的极化属性描述公式,得出各簇丛在不同测度窗口下的受众吸引率、极端受众增长率、极化情感增长率,见表 1。

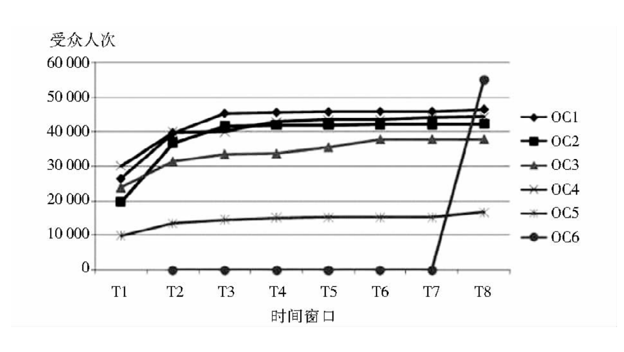


图 5　各测度时段下的观点簇丛极化情感总量

调用测度公式"Epd(TX)""Epd(OC)",得出各观点簇丛在测度时序中的情感极化表现(见表 2),以及各观点簇丛的整体情感极化度(见图 6)。

虽然本次实证数据的应事实体比较集中,但事实上,围绕同一实体的不同话题也呈现出具有差异的极化特征:

通过观察实证结果,笔者发现整个网络舆情场内围绕"霸座男"出现的观点簇丛的情感极性在宏观上具有高度统一性。虽然"占座"和"道歉"两个话题在

表 1　网络舆情场内各簇丛的受众吸引率、极端受众增长率与极化情感增长率

	指标	T1	T2	T3	T4	T5	T6	T7	T8
OC1	Aar		0.22	0.35	0.27	0.42	0.06	0.22	0.36
	Grea		0.46	0.1	0.01	0.005	0.003	0	0.02
	Grpe		0.49	0.14	0.006	0.004	0.002	0	0.01
OC2	Aar		0.32	0.16	0.03	0.01	0.09	0.14	0.19
	Grea		0.79	0.1	0.009	0	0.001	0	0.002
	Grpe		0.87	0.13	0.01	0	0.001	0	0.002
OC3	Aar		0.12	0.14	0.4	0.17	0.35	0.02	0.05
	Grea		0.37	0.02	0.003	0.04	0.08	0	0
	Grpe		0.32	0.06	0.004	0.06	0.06	0	0
OC4	Aar		0.11	0.03	0.17	0.26	0.28	0.55	0.06
	Grea		0.56	0.004	0.09	0.009	0.003	0.02	0.008
	Grpe		0.33	0.005	0.07	0.008	0.002	0.01	0.008
OC5	Aar		0.23	0.32	0.13	0.14	0.22	0.07	0.34
	Grea		0.34	0.07	0.04	0.005	0.05	0	0.09
	Grpe		0.36	0.08	0.04	0.008	0.004	0	0.09
OC6	Aar		–	–	–	–	–	–	0.93
	Grea		–	–	–	–	–	–	0.67
	Grpe		–	–	–	–	–	–	0.48

表 2　网络舆情场内极化观点簇丛在测度时序内的情感极化度

	T1	T2	T3	T4	T5	T6	T7	T8
OC1		0.104 5	0.042	0.002 16	0.001 89	0.000 15	0	0.005 4
OC2		0.265 6	0.018 4	0.000 85	0	0.000 09	0	0.000 38
OC3		0.041 4	0.005 6	0.001 4	0.008 5	0.024 5	0	0
OC4		0.048 95	0.000 135	0.013 6	0.002 21	0.000 7	0.008 25	0.000 48
OC5		0.080 5	0.024	0.005 2	0.000 91	0.005 94	0	0.030 6
OC6		–	–	–	–	–	–	0.534 75

网络舆情场形成初期便引发了广泛的议论,但聚合形成的观点簇丛却并未呈现出很高的情感极化度。而"论文抄袭"和"处罚"两个笔者认为敏感且会引发情感极化的话题,在整个测度时序内也表现平庸。反之,"高学历"和"轮椅挑衅"两个话题呈现出较强的情感极化度。

上述现象的出现主要由以下原因造成:①在高铁霸座事件中,虽然应事实体较为统一,但仍然演化出繁多的舆情子话题,这使得网络舆情场中的受众注意力出现跳跃式转移,各观点簇丛的情感动员普遍缺乏持续性。又由于 OC1/OC2/OC3/OC5 在时间窗口中均出现极化指标为 0 的情况,抵消其前期的良好极化指标表现。②围绕"高学历"形成的观点簇丛虽然在整个测度周期中的极化表现一般,但与其他观点簇丛相比,应事实体身份的话题能比较容易与新进出现的舆情话题对接,所以受到了较持续的关注,情感极化度较高。

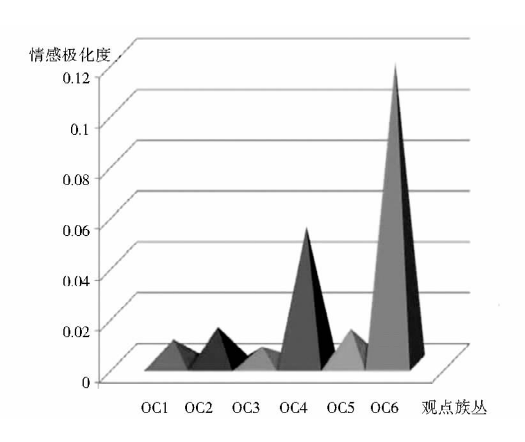


图 6　测度时间内各观点簇丛的整体情感极化度

③应事实体挑衅的事件是在测度时段之末尾出现，所以情感极化度最高。另外，由于此话题的出现，使得原本以进入隐遁期的网络舆情场再次被激活，随时间流逝而淡化的话题再次获得关注，而话题兼容性较好的"高学历"则随着"轮椅挑衅"话题的出现，在最后一个窗口内再次换发极化活力。

实证得出的情感极性程度说明话题既定特质会在情感极化方面起到较高的影响作用，"占座"和"道歉"两观点簇丛的舆情发酵深度不足，"论文抄袭"和"处罚"两个观点簇丛却缺乏舆情发酵广度。围绕"高学历"话题产生的观点簇丛兼具舆情发酵的深度和广度，所以受到持续关注。

这也说明一时一刻的情感极化未必会引发震荡，所以在实证研究中所聚焦的网络舆情场内，"轮椅挑衅"产生的观点簇丛或将产生进一步的情感极化，但其最终演进态势还需要延长测度时序来完成判断。然而受到持续关注的话题却极有可能在整个舆情事件进程中造成不可遇见的后果，在情感极化测度和网络舆情管控实践中必须对此类话题给予足够的重视，而应事实体"座霸男"也应该得到相应的重视，因为在场域内该实体激发起舆情受众情感倾向的高度一致，如不采取措施，易于与其他同质事件相互作用，进而再次引起网络民众情感的集体振荡。

5　结语

笔者提出的对网络舆情场中观点簇丛的情感极化度进行测算的方法，其本质是从 3 个角度综合考量场域内各极化观点簇丛的情感连贯性表现。通过在时间序列内的极化指标的计算，最终得出一个观点簇丛的整体情感极化值，为网络舆情受众极化的干预治理提供量化依据。然而由于时间和能力有限，笔者提出的研究方法尚且存在一定局限：测度必须依托一个连续的时间段，无法仅通过某一时间节点上的极化指标完成对观点簇丛极化程度的整体判断。相信随着此研究的进一步深入，上述问题将会得到妥善的解决。

参考文献：

[1] 高俊峰，宋绍成. 网络舆情信息受众的观点认知距离计量研究[J]. 图书情报工作，2016，60(20)：77－85.
[2] 乐国安，薛婷，陈浩. 网络集群行为的定义和分类框架初探[J]. 中国人民公安大学学报(社会科学版)，2010(6)：99－104.
[3] MCGARTY C，BLIUC A M，THOMAS E F，et al. Collective action as the material expression of opinion-based group membership[J]. Journal of social issues，2010，65(4)：839－857.
[4] YZERBYT V，DUMONT M，WIGBOLDUS D，et al. I feel for us：the impact of categorization and identification on emotions and action tendencies[J]. British journal of social psychology，2011，42(4)：533－549.
[5] BKOWICZ P，SOBKOWICZ A. Dynamics of hate based networks[J]. The European physical journal B，2010，83(4)：633－643.
[6] 石密，刘建准. 网络集体行为意向：概念、测量及形成要素[J]. 情报杂志，2017，36(5)：101－105.
[7] 杜杨沁，霍有光，锁志海. 基于分位数回归的网络群体极化度量[J]. 图书情报工作，2011，55(24)：38－43.
[8] 张润莲，兰月新，王彩华，等. 网络群体性事件演化博弈分析及对策研究[J]. 图书与情报，2016(4)：24－30.
[9] 黄微，宋先智，高俊峰. 网络舆情场中信息受众观点群落的连接鲁棒性测度及实证研究[J]. 情报学报，2017，36(5)：503－510.
[10] 高俊峰. 网络舆情场中信息受众观点群落的凝聚鲁棒性测度研究[J]. 情报杂志，2018，37(4)：106－113.
[11] 吴诗贤，张必兰. 基于观点势场的舆情极化预测模型[J]. 图书情报工作，2015，59(19)：108－112.
[12] 桑斯坦. 网络共和国：网络社会中的民主问题[M]. 黄维明，译. 上海：上海人民出版社，2008.
[13] CHEN H M. Group polarization in virtual communities：the case of stock message boards[J]. Ischools，2013，4(2)：185－195.
[14] 查先进. 信息分析[M]. 武汉：武汉大学出版社，2014.
[15] ZHANG H P，YU H K，XIONG D Y，et al. HHMM-based Chinese lexical analyzer ICTCLAS[C]//Second SIGHAN workshop affiliated with 41th ACL. Stroudsburg：Association for Computational Linguistics，2003：184－187.
[16] 徐琳宏，林鸿飞，潘宇. 情感词汇本体的构造[J]. 情报学报，2008，27(2)：180－185.
[17] 高歌，罗琪玫，王宇. 基于 HNC 理论的文本情感倾向性分析[J]. 数据分析与知识发现，2017，1(8)：85－91.

作者贡献说明：

高俊峰：提出研究思路，撰写论文；
黄微：提出研究思路、设计研究目标与研究方案。

Measuring the Emotional Polarization of Opinion Cluster in Network Public Opinion Field

Gao Junfeng[1] Huang Wei[2]
[1] Beihua University Library, Jilin 132013
[2] Institute of Information Resource Management and Service, Jilin University, Changchun 132013

Abstract: [**Purpose/significance**] This paper proposes a method to measure the emotional polarization degree of the opinion cluster in network public opinion field, which provides a basis for quantifying the emotional situation of opinion audience, identifying strongly polarized groups. [**Method/process**] Firstly, the conditions of emotional polarization of public opinion audience are defined. Then, by setting conditional thresholds, the opinion cluster satisfying the conditions is screened out. Based on that, three polarization descriptive indexes (audience attraction rate/growth rate of extreme audience/growth rate of polarized emotion) are introduced to describe the emotional fermentation degree of opinion clusters in the measurement time phase. Finally, the smoothing weights are introduced to weigh and synthesize the polarization performance of each opinion cluster in the measurement time series, and the overall emotional polarization degree is obtained. [**Result/conclusion**] The emotional performance of opinion clusters at each time phase can provide judgement basis for phased emotional polarization intervention of public opinion audience. The comprehensive emotional polarization value is helpful to accurately identify the sensitive topics and groups in the field, and providing convenience for precise control of network public opinion.

Keywords: network public opinion governance　emotional polarization measurement　opinion analysis　network public opinion field

《图书情报工作》投稿作者学术诚信声明

《图书情报工作》一直秉持发表优秀学术论文成果、促进业界学术交流的使命,并致力于净化学术出版环境,创建良好学术生态。2013 年牵头制订、发布并开始执行《图书馆学期刊关于恪守学术道德净化学术环境的联合声明》(简称《声明》)(见:http://www.lis.ac.cn/CN/column/item202.shtml),随后又牵头制订并发布《中国图书馆学情报学期刊抵制学术不端联合行动计划》(简称《联合行动计划》)(见:http://www.lis.ac.cn/CN/column/item247.shtml)。为贯彻和落实这一理念,本刊郑重声明,即日起,所有投稿作者须承诺:投稿本刊的论文,须遵守以上《声明》及《联合行动计划》,自觉坚守学术道德,坚决抵制学术不端。《图书情报工作》对一切涉嫌抄袭、剽窃等各种学术不端行为的论文实行零容忍,并采取相应的惩戒手段。

《图书情报工作》杂志社